Identifing The Fraud Data In Client Card

Dr. G. Sivakumar[1], Ms. S. Haridharani[2], Ms. J. Jerrine Varshini[3], Ms.P.Keerthana[4]

[1]Assistant Professor, Dept of Computer Science and Engineering

[2, 3, 4]Dept of Computer Science and Engineering

[1, 2, 3, 4] Erode Sengunthar Engineering College, Perundurai, Erode, Tamil Nadu, India.

Abstract- *Credit card fraud has become a major problem in the financial industry, necessitating the implementation of effective fraud detection systems. In this study, we examine the application of three machine learning techniques—Naive Bayes, Logistic Regression, oneR, and the J48 (C4.5) decision tree—for the identification of credit card fraud.The study uses historical credit card transaction data that includes both valid and fraudulent purchases. Following pre-processing and feature engineering, the data is separated into training and testing sets. On the testing set, we train each model and evaluate its performance using a variety of measures such as accuracy, precision, and recall, as well as F1-score and ROC-AUC. The findings demonstrate the efficacy of each algorithm in detecting credit card fraud. Comparisons across the models illustrate their various strengths and shortcomings, offering useful information for determining the best method. Furthermore, we investigate the prospect of increasing detection accuracy using ensemble approaches such as Voting Classifier, Bagging, or Boosting to improve the overall fraud detection system. Finally, the study intends to help financial institutions choose appropriate machine learning algorithms for credit card fraud detection, highlighting the necessity of constant monitoring and model modification in combating emerging fraud strategies efficiently.*

Keywords- C Credit card, machine learning, fraud detection, banking sectors.

I. INTRODUCTION

1.1 CREDIT CARD

Credit cards have transformed how we make purchases and manage our finances. A credit card is a type of plastic payment card that is issued by a bank or other financial institution and enables the cardholder to borrow money for the purpose of making purchases of goods and services.Unlike debit cards, which remove cash straight from the cardholder's bank account, credit cards provide a line of credit that must be returned at a later date, typically with interest. The cardholder may use the credit card for a variety of purchases, both in person and online, making it a convenient and generally recognized payment method across the world. Credit cards have become an essential aspect of modern consumer culture, giving people more purchasing power and the freedom to manage their finances properly. When a consumer applies for a credit card and is authorized, the financial institution sets them a credit limit, which is the most money they may borrow with the card. The cardholder may use the credit card to make purchases up to the limit. Each month, the credit card firm delivers a billing statement outlining the transactions that occurred during that time period as well as the minimum amount owing. While paying the minimal amount keeps the account in good standing, it is best to pay off the whole debt to prevent interest costs. If the cardholder fails to make the minimum payment on time, they risk incurring late penalties and harming their credit score. Credit cards sometimes come with a variety of incentives and advantages, such as cash back, travel points, or discounts, to encourage card use and loyalty.

1.2 FRAUD DETECTION

Fraud detection is a vital use of data analysis and machine learning that aims to identify and prevent fraudulent activity and transactions across several domains. As technology progresses and financial transactions move to digital platforms, the danger of fraudulent activity increases, making fraud detection more important than ever. Whether it's credit card fraud, insurance fraud, identity theft, or internet scams, companies and financial institutions use sophisticated fraud detection systems to protect their assets, consumers, and preserve faith in their services. Fraud detection systems examine massive volumes of transactional and behavioural data to find anomalies and suspect trends. Fraud may have serious consequences, including financial losses, tarnished reputations, and impaired consumer trust.

II. LITERATURE REVIEW

2.1 UNCERTAINTY- AWARE CREDIT CARD FRAUD DETECTION USING DEEP LEARNING

Maryam Habibpour [1] and colleagues propose in this work Numerous research have used deep neural networks (DNNs) to identify credit card fraud, with the goal of improving point prediction accuracy and avoiding unwanted biases through the development of various network architectures or learning models. It is critical to measure uncertainty in conjunction with point estimate because it lowers model unfairness and allows practitioners to build dependable systems that prevent making incorrect judgments due to uncertainty. Because fraudsters continuously change their strategies, DNNs meet observations that do not come from the same process as the training distribution. Furthermore, because to the time-consuming nature of the procedure, only few transactions are reviewed by experienced specialists in order to update DNNs. These characteristics make it necessary to clearly evaluate the uncertainty associated with DNN predictions in real-world card fraud detection scenarios.

2.2 CREDIT CARD FRAUD DETECTION IN THE ERA OF DISRUPTIVE TECHNOLOGIES: A SYSTEMATIC REVIEW

Credit card fraud is becoming an increasingly serious issue with the advent of new technologies and communication channels, such as contactless payment. This article provides a comprehensive review of current research on detecting and forecasting fraudulent credit card transactions from 2015 to 2021. We analyze and classify 40 selected papers based on the topics they address, including the class imbalance problem and feature engineering, as well as the machine learning approaches they employ, encompassing both conventional and deep learning models. Our analysis reveals that deep learning has been underexplored, suggesting a need for more research to tackle the challenges in detecting credit card fraud using advanced technologies such as big data analytics, large-scale machine learning, and cloud computing. This study serves as a valuable resource for both academic and industry researchers in evaluating financial fraud detection systems and developing reliable solutions by highlighting current research challenges and future research opportunities.

2.3 MACHINE LEARNING BASED ON RESAMPLING APPROACHES AND DEEP REINFORCEMENT LEARNING FOR CREDIT CARD DETECTION SYSTEMS

Dr. Tran Khanh Dang and colleagues emphasize that unbalanced datasets are a fundamental concern in constructing reliable credit card fraud (CCF) detection systems. In their paper, they study and evaluate the latest advancements in deep reinforcement learning (DRL) and machine learning (ML) algorithms for CCF detection, covering both fraud and non-fraud labels. To address dataset imbalance, they employ SMOTE and ADASYN resampling methods. The balanced dataset generated from these techniques is then used to train ML algorithms for creating CCF detection models. Additionally, they use the original imbalanced CCF dataset to develop detection algorithms with DRL.

2.4 ON THE BLACK-BOX CHALLENGE FOR FRAUD DETECTION USING MACHINE LEARNING (II) NONLINEAR ANALYSIS THROUGH INTERPRETABLE AUTOENCODERS

Jacobo Chaquet-Ulldemolins [4] et al. proposed this system. Artificial intelligence (AI) has lately gained popularity in the global economy due to its exceptional ability to analyse and model data in a variety of disciplines. As a result of this condition, society is rapidly becoming more automated, and these new approaches may be combined to form a beneficial tool for addressing the difficult challenge of credit fraud detection. However, stringent restrictions prevent financial institutions from adhering to them while utilizing modern procedures. Methodologically, autoencoders have proven effective in identifying nonlinear features across various problem domains. Nonetheless, autoencoders are often considered "black boxes" due to their opacity. In this study, we present an interpretable and unbiased approach to credit card fraud detection (CFD).

2.5 CREDIT CARD FRAUD DETECTION USING A NEW HYBRID MACHINE LEARNING ARCHITECTURE

Esraa Faisal Malik [5] highlights the increasing impact of financial crimes on financial institutions. Various single and hybrid machine learning algorithms have been employed to detect crimes such as credit card fraud. However, these techniques face significant limitations due to a lack of additional research on alternative hybrid algorithms for specific datasets. This paper proposes and tests seven hybrid machine learning models for detecting fraudulent activities using a real-world dataset. Initially, modern machine learning techniques were applied to detect credit card fraud, and the best-performing single algorithm from this phase was used to develop the hybrid approaches. The hybrid models were evaluated in two phases. Our findings indicate that the hybrid model combining AdaBoost and LGBM outperforms the others. Future research should focus on exploring different hybridization strategies and algorithms within the credit card fraud domain.

III. RELATED WORK

In today's digital economy, credit cards have become essential, and their usage has significantly increased, resulting in a rise in credit card theft. Machine learning (ML) algorithms have been employed to detect credit card fraud, but the dynamic shopping habits of cardholders and class imbalance issues make it challenging for ML classifiers to perform optimally. This paper introduces a robust deep learning method to address these challenges. It incorporates a multilayer perceptron (MLP) as the meta-learner and utilizes long short-term memory (LSTM) and gated recurrent unit (GRU) neural networks as base learners within a stacking ensemble architecture. The edited nearest neighbour (SMOTE-ENN) method and hybrid synthetic minority oversampling technique are used to balance the class distribution in the dataset. According to experimental data, the suggested deep learning ensemble achieves sensitivity and specificity values of 1.000 and 0.997, respectively, when paired with the SMOTE-ENN approach. These outcomes outperform those of other widely applied machine learning algorithms and classifiers in the literature.

IV. METHODOLOGY

The historical credit card transaction data collection, The data pre-processing step entails gathering historical credit card transaction data and undertaking extensive data cleaning and feature engineering. To create a sizable dataset suitable for training the models, pertinent data is obtained, including transaction amount, location, time of day, and cardholder behaviour.

To improve the accuracy and resilience of the fraud detection system, we use ensemble modeling approaches. This entails merging the results of multiple machine learning algorithms, such as Naive Bayes, Logistic Regression, oneR, and the J48 (C4.5) decision tree, utilizing methods like Voting Classifier or Bagging. By combining predictions from various models, we hope to decrease false positives and false negatives, resulting in more reliable and efficient fraud detection. The suggested system's fundamental function is real-time fraud detection. Once trained and refined, the ensemble model is deployed in a real-time credit card transaction processing system. As new transactions occur, the model quickly analyses their risk level and issues alerts for possible fraudulent activity. The system's real-time nature allows for quick replies, reducing possible losses and providing a safe experience for cardholders.

A. Load Data

The goal of the feature selection module is to choose the pre-processed data's most pertinent and instructive features. To determine the most important characteristics that contribute to the detection of credit card fraud, it makes use of a variety of methodologies, including statistical testing, correlation analysis, or feature importance rankings. The credit card transaction data must first be loaded into a framework for machine learning. Several sources, including financial institutions, credit card firms, and publicly accessible databases, can provide the data.

B. Data Pre-Processing

$$LR(z) = \frac{1}{1+e^{z}}$$

The initial cleaning and preparing of the credit card transaction data is handled by the Data Pre-processing module. It carries out operations such handling missing values, identifying and treating outliers, and normalizing or scaling data. It also finds and fixes any problems with the quality of the data that can affect how accurate the future models are.

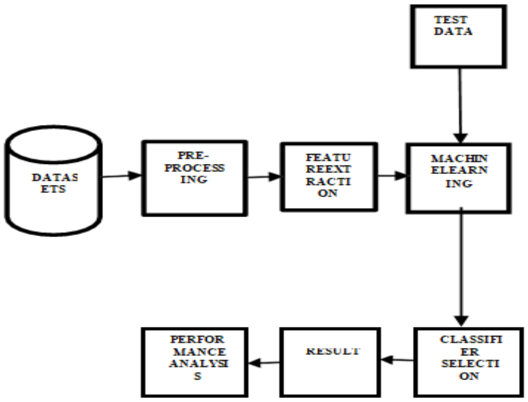

Figure 1. Block diagram

V. ALGORITHM DETAILS

A. Naïve Bayes Algorithm

A straightforward probabilistic algorithm used in classification, the Naïve Bayes algorithm determines its probability value by calculating value combinations and frequency combinations from the related collection. According to this algorithm, every attribute is considered independent. Naïve Bayes classification requires multiple cues or instructions to identify the class of the data to be analysed. Consequently, Equation 1 is used.

$$P(C|F_1 \ldots F_n) = \frac{P(C).P(F_1 \ldots F_n|C)}{P(F_1 \ldots F_n)}$$

Variable C in Equation 1 represents a class, while variables F1...Fn stand for attributes needed to do a classification. As a result, the posterior probability of matching data to a particular feature in the C class is equal to the probability of the sample characteristics emerging in the C class multiplied by the likelihood of those characteristics in the class, divided by the probability of the sample characteristics globally (evidence).

Input:

Training dataset T,

F= (f1, f2, f3,.., fn) // value of the predictor variable in testing dataset.
Output:

A class of testing dataset. Step:

1. Read the training dataset T;
2. Calculate the mean and standard deviation of the predictor variables in each class;
3. Repeat

   Calculate the probability of *fi* using the gauss density equation in each class;
   Until the probability of all predictor variables (f1, f2, f3,.., fn) has been calculated.
4. Calculate the likelihood for each class;
5. Get the greatest likelihood;

B. Logistic Regression

Logistic regression is basically a supervised classification algorithm. In a classification problem, the target variable (or output), y, can take only discrete values for given set of features (or inputs), X.

When a decision threshold is included, logistic regression turns into a classification method. A crucial component of logistic regression is threshold setting, which is contingent on the nature of the classification problem. The optimal threshold value for logistic regression, based on feature selection applied to this data set, is 0.6.

C. J48 decision tree

The target value from a dataset is analysed using a prediction approach based on many specified attributes. It extracts the feature that distinguishes multiple occurrences from the training set. To reach the highest information acquisition, these cases are categorised even more. This process is repeatedly applied to the smaller subsets until every occurrence is correctly classified in its class. As seen in Figure 1, the top level consists of a single header node that points to its offspring. Internal nodes represent attributes, while branches indicate potential values for these attributes.

VI. CONCLUSION

The credit card fraud detection system, which uses Naïve Bayes, Logistic Regression, and J48 (C4.5) algorithms, provides a comprehensive and robust strategy to prevent fraudulent financial transactions. The system's goal is to identify fraudulent transactions with high accuracy, precision, recall, and F1-score while minimizing false positives by meticulous data pre-processing, feature selection, and ensemble modeling. If a real- time fraud detection module is added, it improves the system's efficacy by immediately raising alarms for possibly fraudulent transactions, allowing for early action to avert losses. The user-friendly interface allows customers to configure the system's parameters, allowing for specialized fraud detection depending on unique needs. With continual monitoring and updates, the system can react to developing fraud tendencies, ensuring a high degree of security in an ever-changing financial environment.

REFERENCES

[1] M. Habibpour, H. Gharoun, M. Mehdipour, A. Tajally, H. Asgharnezhad, A. Shamsi, A. Khosravi, M. Shafie-Khah, S. Nahavandi, and J. P. S. Catalao, "Uncertainty-aware credit card fraud detection using deep learning 2021; arXiv:2107.13508.

[2] A. Cherif, A. Badhib, H. Ammar, S. Alshehri, M. Kalkatawi, and A. Imine. "Credit card fraud detection in the era of disruptive technologies: A systematic review." J. King Saud Univ. Computer and Information Science, vol. 35, no. 1, pp. 145- 174, Jan. 2023, doi:10.1016/j.jksuci.2022.11.008.

[3] T. K. Dang, T. C. Tran, L. M. Tuan, and M. V. Tiep. "Machine learning based on resampling approaches and deep reinforcement learning for credit card fraud detection systems." Appl. Sci., vol. 11, no. 21, p. 10004, Oct. 2021; doi: 10.3390/app112110004.

[4] Chaquet-Ulldemolins et al., "On the black-box problem for fraud detection using machine learning (I): Linear models and informative feature selection," Applied Sciences, vol. 12, no. 7, p. 3328, March 2022, doi: 10.3390/app12073328.

[5] E. F. Malik, K. W. Khaw, B. Belaton, W. P. Wong, and X. Chew. "Credit card fraud detection using a new hybrid machine learning architecture." Mathematics, vol. 10, no. 9, p. 1480, April 2022; doi: 10.3390/math10091480.

[6] I. Benchaji, S. Douzi, B. El Ouahidi, and J. Jaafari, "Enhanced credit card fraud detection using attention mechanism and LSTM deep model," J. Big Data, vol. 8, no. 1, p. 151, December 2021; doi: 10.1186/s40537-021-00541-8.

[7] E. Esenogho, I. D. Mienye, T. G. Swart, K. Aruleba, and G. Obaido. "A neural network ensemble with feature engineering for improved credit card fraud detection." IEEE Access, vol. 10, pp. 16400–16407, 2022; doi: 10.1109/ACCESS.2022.3148298.

[8] E. Btoush, X. Zhou, R. Gururaian, K. Chan, and X. Tao, "A survey on credit card fraud detection approaches in banking sector for cyber security," in Proc. 8th International Conf. Behav. Social Comput. (BESC), Oct. 2021, pp. 1-7, doi: 10.1109/BESC53957.2021.9635559.

[9] Y. Xie, G. Liu, C. Yan, C. Jiang, M. Zhou, and M. Li, "Learning transactional behavioral representations for credit card fraud detection IEEE Trans. Neural Networks and Learning Systems, early access, October 5, 2022, doi: 10.1109/TNNLS.2022.3208967.

[10] J. Yang & J. Guan "A heart disease prediction model based on feature optimization and the smote-Xgboost algorithm," Information, vol. 13, no. 10, p. 475, Oct. 2022, doi: 10.3390/info13100475.